SEQUENCE LISTIN

(1) GENERAL INFORMATION:

(i) APPLICANT:

Donson, Jon

Dawson, William O. Grantham, George L.

Turpen, Thomas H. Turpen, Ann Myers Garger, Stephen J. Grill, Laurence K.

- (ii) TITLE OF INVENTION: RECOMBINANT PLANT VIRAL NUCLEIC ACIDS
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Limbach & Limbach
 - (8) STREET: 2001 Ferry Building
 - (C) CITY: San Francisco
 - (D) STATE: CAL
 - (F) ZIP: 94111
- (V) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Patent in Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 923,692
 - (B) FILING DATE: 31-JUL-1992
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 600,244
 - (B) FILING DATE: 22-OCT-1990
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 641,617
 - (B) FILING DATE: 16-JAN-1991
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 310,881
 - (B) FILING DATE: 17-FEB-1989
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,766
 - (B) FILING DATE: 26-FEB-1988
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,771
 - (B) FILING DATE: 26-FEB-1988
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 347,637
 - (B) FILING DATE: 05-MAY-1989
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 363,138
 - (B) FILING DATE: 08-JUN-1989

RECEIVED
JUL 0 8 1993
GROUP 1800

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioned of Patents and Trademarks.

Washington, DC 20231 on.

TIMBACH & LIMBACH

___ By_

(vii) PRIOR APPLICATION DATA: (A) APPLICATION NUMBER: US 219,279 (B) FILING DATE: 15-JUL-1988 (viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Halluin, Albert P, (B) REGISTRATION NUMBER: 28,957 (C) REFERENCE/DOCKET NUMBER: BIOG-20121 USA (ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: 415-433-4150 (B) TELEFAX: 415-433-8716 (2) INFORMATION FOR SEQ ID NO: 1: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 4 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: peptide (iii) HYPOTHETICAL: NO (iv) ANTI-SENSE: NO (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1: Pro Xaa Gly Pro (2) INFORMATION FOR SEQ ID NO: 2: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 13 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (iii) HYPOTHETICAL: NO (iv) ANTI-SENSE: NO (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2: GGGTACCTGG GCC

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 886 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

13

				•				
(ii)	MOLECULE	TYPE: DNA	(genomi	c)				
(iii)	HYPOTHETI	CAL: NO						
(iv)	ANTI-SENS	SE: NO						
(vi)	ORIGINAL (A) ORGA	SOURCE: ANISM: Chir	nëse cuc	cumber				
(vii)	IMMEDIATE	SOURCE: IE: alpha-t	richosa	nthin				
(ix)		E/KEY: CDS ATION: 8		CATION: 8	3877			
(xi),	SEQUENCE	DESCRIPTIO	ON: SEQ	ID-NO: 3	3 :			
CTCGAGG	ATG ATC AGA	TTC TTA G	этс стс	TCT TTG	CTA ATT	CTC ACC	стс	49
	Met Ile Arg	Phe Leu V 5	/al Leu	Ser Leu	Leu Ile 10	Leu Thr	Leu	
TTC CTA	ACA ACT CCT	CT GTG G	AG GGC	GAT GTT	AGC TTC	CGT TTA	TCA	97
Phe Leu 15	Thr Thr Pro	Ala Val G 20	Slu Gly	Asp Val 25	Ser Phe	Arg Leu	Ser 30	
GGT GCA	ACA AGC AG1	TCC TAT G	GA GTT	TTC ATT	TCA AAT	CTG AGA	AAA	145
Gly Ala	Thr Ser Ser	•	Gly Val	Phe Ile 40	Ser Asn	Leu Arg	•	
GCT CTT	CCA AAT GAA	AGG AAA C	TG TAC	GAT ATC	CCT CTG	TTA CGT	TCC	193
Ala Leu	Pro Asn Glu 50	J Arg Lys L	eu Tyr. 55	Asp Ile	Pro Leu	Leu Arg 60	Ser.	
тст стт	CCA GGT TC	CAA CGC T	TAC GCA	TTG ATC	CAT CTC	ACA AAT	TAC	241
Ser Leu	Pro Gly Ser	Gln Arg T	Tyr Ala 70	Leu Ile	His Leu 75	Thr Asr	Tyr	
GCC GAT	GAA ACC ATT	T TCA GTG G	CC ATA	GAC GTA	ACG AAC	GTC TAT	ATT	289
Ala Asp 80	Glu Thr Ile	e Ser Val A 85	lla Ile	Asp Val	Thr Asn 90	Val Tyr	Ile	
ATG GGA	TAT CGC GCT	GGC GAT A	ACA TCC	TAT TTT	TTC AAC	GAG GCT	тст	337
Met Gly 95	Tyr Arg Ala	a Gly Asp 1 100	Thr Ser	Tyr Phe 105	Phe Asn	Glu Ala	Ser 110	
GCA ACA	GAA GCT GCA	A AAA TAT G	STA TTC	AAA GAC	GCT ATG	CGA AAA	GTT	385
Ala Thr	Glu Ala Ala 115		/al Phe	Lys Asp 120	Ala Met	Arg Lys		
ACG CTT	CCA TAT TC	GGC AAT 1	TAC GAA	AGG CTT	CAA ACT	GCT GCC	GGC	433

Thr	Leu	Pro	Tyr 130	Ser	Gly	Asn	Tyr	Glu 135	Arg	Leu	Gln	Thr	Ala 140	Ala	Gly	
AAA	ATA	AGG	GAA	AAT	ATT	CCG	CTT	GGA	стс	CCA	GCT	TTG	GAC	AGT	GCC	481
Lys	lle	Arg 145	Glu	Asn	Ile	Pro	Leu 150	Gly	Leu	Pro	Ala	Leu 155	Asp	Ser	Ala	
ATT	ACC	ACT	TTG	TTT	TAC	TAC	AAC	GCC	AAT	тст	GCT	GCG	TCG	GCA	СТТ	529
Ile	Thr 160	Thr	Leu	Phe	Туг	Tyr 165	Asn	Ala	Asn	Ser	Ala 170	Ala	Ser	Ala	Leu	
ATG	GTA	СТС	ATT	CAG	TCG	ACG	тст	GAG	GCT	GCG	AGG	TAT	AAA	TTT	ATT	577
Met 175	Val	Leu	Ile	Gln	Ser 180	Thr.	Ser	Glu	Ala 185	Ala	Arg	Туг	Lys	Phe	Ile 190	
GAG	CAA	CAA	ATT	GGG	AAG	CGC	GTT	GAC	AAA	ACC	TTC	CTA	CCA	AGT	TTA	625
Glu	Gln	Gln	Ile	Gly 195	Lys	Arg	Val	Asp	Lys 200	Thr	Phe	Leu	Pro	Ser 205	Leu	
GCA	ATT	ATA	AGT	TTG	GAA	AAT	AGT	TGG	тст	GCT	СТС	TCC	AAG	CAA	ATT	673
Ala	Ile	Ile	Ser 210	Leu	Glu	Asn	Ser	Trp 215	Ser	Ala	Leu	Ser	Lys 220	Gln	Ile	
CAG	ATA	GCG	AGT	ACT	AAT	AAT	GGA	CAG	TTT	GAA	ACT	ССТ	GTT	GTG	СТТ	721
Gln	Ile	Ala 225	Ser	Thr	Asn	Asn	Gly 230	Gln	Phe	Glu	Thr	Pro 235	Val	Val	Leu	
ATA	AAT	GCT	CAA	AAC	CAA	CGA'	GTC	ATG	ATA	ACC	AAT	GTT	GAT	GCT	GGA	769
Ile	Asn 240	Ala	Gln	Asn		Arg 245		Met	Ile	Thr	A sn 250	Val	Asp	Ala	Gly	
GTT	GTA	ACC	TCC	AAC	ATC	GCG	TTG	CTG	CTG	AAT	CGA	AAC	AAT	ATG	GCA	817
Val 255	Val	Thr	Ser	Asn	Ile 260	Ala	Leu	Leu	Leu	Asn 265	Arg	Asn	Asn	Met	Ala 270	
GCC	ATG	GAT	GAC	GAT	GTT	ССТ	ATG	ACA	CAG	AGC	TTT	GGA	TGT	GGA	AGT	865
Ala	Met	Asp	Asp	Asp 275	Val	Pro	Met	Thr	Gln 280	Ser	Phe	Gly	Cys	Gly 285	Ser	
TAT	GCT	ATT	TAG	[AAC]	rcg /	\G										886
Туг	Ala	Ile	290)		٠										

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 289 amino acids

(B) TYPE: amino acid

		(3) 1	YPE:	H: 4 : ami	ino a	cid	aci	ds						
(ii	i)	MO	DLECU	JLE 1	YPE:	pro	oteir	1							
(xi	i)	SE	QUEN	ICE (DESC	RIPTI	ON:	SEQ	ID N	10: 6	5:				
Met 1	Gln	Val	Leu	Asn 5	Thr	Met _.	Val	Asn	Lys 10	His	Phe	Leu	Ser	Leu 15	Ser
Val	Leu	Ile	Val 20	Leu	Leu	Gly	Leu	Ser 25	Ser	Asn	Leu	Thr	Ala 30	Gly	Gln
Val	Leu	Phe 35	Gln	Gly	Phe	Asn	Trp 40	Glu	Ser	Trp	Lys	Glu 45	Asn	Gly	Gly
Trp	Туг 50	Asn	Phe	Leu	Met	Gly 55	Lys	Val	Asp	Asp	1 l e 60	Ala	Ala	Ala	Gly
Ile 65	Thr	His	Val	Trp	Leu 70	Pro	Pro	Pro	Ser	His 75	Ser	Val	Gly	Glu	Gln 80
Gly	Туг	Met	Pro	Gly 85	Arg	Leu	Туг	Asp	Leu 90	Asp	Ala	Ser	Lys	Туг 95	Gly
Asn	Glu	Ala	Gln 100	Leu	Lys	Ser	Leu	I l e 105	Glu	Ala	Phe	His	Gly 110	Lys	Gly
Val	Gln	Val 115	Ιļe	Ala	Asp	Ile	Val 120	1 le	Asn	His	Arg	Thr 125	Ala	Glu	His
Lys	Asp 130	Gly	Arg	Gly	Ile	Tyr 135	Cys	Leu	Phe	Glu	Gly 140	Gly	Thr	Pro	Asp
Ser 145	Arg	Leu	Asp	Trp	Gly 150	Pro	His	Met	Ile	Cys 155	Arg	Asp	Asp	Pro	Tyr 160
Gly	Asp	Gly	Thr	Gly 165	Asn	Pro	Asp	Thr	Gly 170	Ala	Asp	Phe	Ala	Ala 175	Ala
Pro	Asp	Ile	Asp 180	His	Leu	Asn	Lys	Arg 185	Val	Gln	Arg	Glu	Leu 190	Ile	Gly
Тгр	Leu	Asp 195	Trp	Leu	Lys	Met [.]	Asp 200	Ile	Gly	Phe	Asp	Ala 205	Тгр	Arg	Leu
Asp	Phe 210	Ala	Lys	Gly	Tyr	Ser 215	Ala	Asp	Met	Ala	Lys 220	Ile	Туг	Ile	Asp
Ala 225	Thr	Glu	Pro	Ser	Phe 230	Ala	Val	Ala	Glu	I l e 235	Trp	Thr	Ser	Met	Ala 240
Asn	Gly	Gly	Asp	Gly 245	Lys	Pro	Asn	Tyr	Asp 250		Asn	Ala	His	Arg 255	Gln
Glu	Leu	Val	Asn	Trp	Val	Asp	Arg	Val	Gly	Gly	Ala	Asn	Ser	Asn	Gly

Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val Ala Val Glu

Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala Pro Gly Met 295 . 300 Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp Asn His Asp 310 315 Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp Lys Val Met 325 330 Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro Cys Ile Phe 345 Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile Glu Arg Leu 360 Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser Glu Leu Arg . 370 375 380 Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile Asp Gly Lys 390 395 Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His Leu Ile Pro 410 Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala Ile Trp Glu 425 Lys Ile (2) INFORMATION FOR SEQ ID NO:7: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 709 base pairs (B) TYPE: nucleic acid (G) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA to mRNA (iii) HYPOTHETICAL: NO ANTI-SENSE: NO (iv) (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (vii) IMMEDIATE SOURCE: (B) CLONE: alpha-hemoglobin (ix) (A) NAME/KEY: transit_peptide (B) LOCATION: 26. .241 (B) LOCATION: 26. .241 (ix) FEATURE: (A) NAME/KEY: CDS

(B) LOCATION: 245. .670

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CTC	GAGG	GCA 1	rctg/	ATCTI	T C	AAGA/	ATGG	CACA	NAAT1	TAAC	AACA	TGG	CAC /	AAGGG	SATACA	60
AAC	CTT	AT (CCAA	ATTC	CA AT	ттсс	CATA	A ACC	CCAA	AGTT	CCTA	AATO	:TT (CAAGT	тттст	120
TGT	TTTC	GA- 1	GTA/	\AAA/	C TO	GAAA	ATTC	AGC	NAA†1	CT A	ATGTI	GGT1	T TO	GAAAA	AAGA	180
TTC	ATT:	TTT #	ATGC	\AAA(ST T1	TTGTT	CCT	TAC	GATI	TCA	GCAC	GTG	STA (GAGTI	TCTTG	240
CAT	G GT(СТС	S TC1	r cc1	GCC	C GAC	CAAC	G ACC	CAAC	GTO	CAAC	GCC	GC	C TGG	GGC	289
		l Lei	ı Ser	Pro	Ala 5	•) Lys	s Thi	- Asr	val	. •	a Ala	a Ala	a Trp	Cly 15	
AAG	GTT	GGC	GCG	CAC	GCT	GGC	GAG	TAT	GGT	GCG	GAG	GCC	CTG	GAG	AGG	337
Lys	Val	Gly	Ala	His 20	Ala	Gly	Glu	Туг	Gly 25	Ala	Glu	Ala	Leu	Glu 30	Arg	
ATG	TTC	CTG	TCC	TTC	ccc	ACC _.	ACC	AAG	ACC	TAC	TTC	CCG	CAC	ттс	GAC	385
Met	Phe	Leu	Ser 35	Phe	Pro	Thr	Thr	Lys 40	Thr	Tyr	Phe	Рго	His 45	Phe	Asp	
CTG	AGC	CAC	GGC	тст	GCC	CAG	GTT	AAG	GGC	CAC	GGC	AAG	AAG	GTG	GCC	433
Leu	Ser	His 50	Gly	Ser	Ala	Gln	Val 55	Lys	Gly	His	Gly	Lys 60	Lys	Val	Ala `	
GAC	GCG	CTG	ACC	AAC	GCC	GTG	GCG	CAC	GTG	GAC	GAC	ATG	CCC	AAC	GCG	481
Asp	Ala 65	Leu	Thr	Asn	Ala	Val 70	Ala	His	Val	Asp	Asp 75	Met	Pro	Asn	Ala	
CTG	TCC	GCC	CTG	AGC	GAC	CTG	CAC	GCG	CAC	AAG	CTT	CGG	GTG	GAC	CCG	529
Leu 80	Ser	Ala	Leu	Ser	Asp 85	Leu	His	Ala	His	Lys 90	Leu	Arg	Val	Asp	Pro 95	
GTC	AAC	TTC	AAG	СТС	CTA	AGC	CAC	TGC	CTG	CTG	GTG	ACC	CTG	GCC	GCC	577
Val	Asn	Phe	Lys	Leu 100	Leu	Ser	His	Cys	Leu 105	Leu	Val	Thr	Leu	Ala 110	Ala	
CAC	СТС	ccc	GCC	GAG	TTC	ACC	ССТ	GCG	GTG	CAC	GCC	TCC	CTG	GAC	AAG	625
His	Leu	Pro	Ala 115.	Glu	Phe	Thr	Pro	Ala 120	Val	His	Ala	Ser	Leu 125	Asp	Lys	
TTC	CTG	GCT	тст	GTG	AGC	ACC	GTG	CTG	ACC	TCC	AAA	TAC	CGT	TAAG	CTGGAG	677
Phe	Leu	Ala 130	Ser	Val	Ser	Thr	Val 135	Leu	Thr	Ser	Lys	Tyr 140	Arg			

(i)	(<i>I</i>	A) (B) 1	ENG1		141 nino	ami	ino a	acids	3					
(ii)	MC	DLECL	JLE 1	TYPE:	pro	oteir	1				•			
(i)	()	SE	EQUE	ICE (ESCF	RIPTI	ON:	SEQ	ID I	10:8:	:				
al 1	Leu	Ser	Pro	Ala 5	Asp	Lys	Thr	Asn	Val 10	Lys	Ala	Ala	Trp	Gly 15	Lys
al	Gly	Ala	His 20	Ala	Gly	Glư	Tyr	Gly 25	Ala	Glu	Ala	Leu	Glu 3 0	Arg	Met
he	Leu	Ser 35	Phe	Pro	Thr	Thr	Lys 40	Thr	Tyŕ	Phe	Pro	His 45	Phe	Asp	Leu
er	His 50	Gly	Ser	Ala	Gln	Val 55	Lys	Gly	His	Gly	Lys 60	Lys	Val	Ala	Asp
la 65	Leu	Thr	Asn	Ala	Val 70	Ala	His	Val	Asp	Asp . 75	Met	Pro	Asn	Ala	Leu 80
er	Ala	Leu	Ser	Asp 85	Leu	His	Ala	His	Lys 90	Leu	Arg	Val	Asp	Pro 95	Val
sn	Phe	Lys	Leu 100	Leu	Ser	His	Cys	Leu 105	Leu	Val	Thr	Leu	Ala 110	Ala	His
eu	Pro	Ala 115	Glu	Phe	Thr	Pro	Ala 120	Val	His	Ala	Ser	Leu 125	Asp	Lys	Phe
eu	Ala 130	Ser	Val	Ser	Thr	Val 135	Leu	Thr	Ser	Lys	Tyr 140	Arg			
2)	INFO	ORMAT	T I ON	FOR	SEQ	ID N	10:9:	•							
(i)	() ()	A) LE B) TY C) S1	ENGTI (PE: [RAN(CHARA I: 74 nucl DEDNE DGY:	3 ba leic SS:	ase p acid	oairs d							
(ii)	MC	DLECI	JLE '	TYPE:	: cDN	IA to	omRI	IA						
iii)	н	POTI	IET I	CAL:	NO ·									,
(iv	()	Al	NT [- S	SENSI	E: NO)									
(vi)				SOUR(no sa	ap i er	าร				٠		
vii)				sour		nemog	globi	in			•			

(ix)

FEATURE:

(ix)	FEATURE		•							
	• •	AME/KEY: OCATION:	CDS 24568	35						
_										
(xi)	SEQUENC	CE DESCR	IPTION:	SEQ II) NO: 9	?:				
CTCGAGGGG	A TCTGAT	TCTTT CA	AGAATGG	C ACAA	ATTAAC	AACATO	GCAC /	AAGGG	ATACA	60
AACCCTTAA	T CCCAAT	TTCCA AT	TTCCATA	A ACCC	CAAGTT	CCTAAA	TCTT	CAAGT	тттст	120
TGTTTTTGG	A TCTAA	AAAAC TG	AAAAATTI	CAGCA	AATTCT	ATGTTO	GTTT	rgaaa	AAAGA	180
TTCAATTTT	T ATGCA/	AAAGT TT	TGTTCCT	T TAGG	ATTTCA	GCAGG1	GGTA (GAGTT	TCTTG	240
GATG GTG	CAC CTG	ACT CCT	GAG GA	G AAG	TCT GC	GTT A	CT GC	СТО	TGG	289
Val 1	His Leu	Thr Pro	Glu Gli	u Lys :	Sër Ala 10		hr Ala	a Leu	1 Trp	
	TO 440				OT 040			400	070	777
GGC AAG G	IIG AAC I	UIG GAI	GAA GII	GG1 G	GI GAG	GCC C1	6 666	AGG	CIG	337
Gly Lys V	al Asn \	Val Asp 20	Glu Val		ly Glu 25	Ala Le	u Gly	Arg 30	Leu	
CTG GTG G	TC TAC	CCT TGG	ACC CAG	AGG T	TC TTT	GAG TO	C TTT	GGG	GAT	385
Leu Val V	al Tyr i 35	Pro Trp	Thr Gln	Arg Pl	he Phe	Glu Se	r Phe 45	Gly	Asp	
CTG TCC A	CT CCT (GAT GCT	GTT ATG	GGC A	AC CCT	AAG G1	G AAG	GCT	CAT	433
Leu Ser T	hr Pro <i>l</i> 50	Asp Ala	Val Met 55	Gly A	sn Pro		ıl Lys 0	Ala	His	
GGC AAG A	AA GTG (CTG GGT	GCC TTT	AGT G	AT GGC	CTG GC	T CAC	CTG	GAC	481
Gly Lys·L 65	ys Val I	Leu Gly	Ala Phe 70	Ser A	sp Gly	Leu Al 75	a His	Leu	Asp	
AAC CTC A	AG GGC	ACC TTT	GCC, ACC	CTG A	GT GAG	CTG CA	C TGT	GAC	AAG	529
Asn Leu L 80	ys Gly	Thr Phe 85	Ala Thr	Leu S	er Glu - 90	Leu Hi	s Cys	Asp	Lys 95	
CTG CAC G	TG GAT (CCT GAG	AGC TTC	AGG C	TC CTA	GGC AA	C GTG	CTG	GTC	577
Leu His V	•	Pro Glu 100	Ser Phe	-	eu Leu 05	Gly As	n Val	Leu 110	Val	
TGT GTG C	TG GCG (CAT CAC	TTT GGC	AAA G	AA TTC	ACC CO	A CCA	GTG	CAG	625
Cys Val L	eu Ala (His His	Phe Gly	Lys G 120	lu Phe	Thr Pi	o Pro 125	Val	Gln	
GCT GCC T	AT CAG	AAA GTG	GTG GCT	GGT G	TG GCT	AAT GO	C CTG	GCC	CAC	673
Ala Ala T	yr Gln I	Lys Val	Val Ala	Gly V	al Ala	Asn Al	a Leu	Ala	His	

(A) NAME/KEY: transit_peptide (B) LOCATION: 26. .241

(B) LOCATION: 26..241

130 135 14

AAG TAT CAC TAAGCTCGCT TTCTTGCTGT CCAATTTCTA TTAAAGGTTC

Lys Tyr His 145

CTTTGTGGGG TCGAGGTCGA C

743

722

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 146 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp Gly

1 5 10 15

Lys Val Asn Val Asp Glu Val Gly Glu Ala Leu Gly Arg Leu Leu 20 25 30

Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp Leu
35 40 45

Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His Gly
50 55 60

Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp Asn 65 70 75 80

Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys Leu 85 90 95

His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val Cys 100 105 110

Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln Ala 115 120 125

Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His Lys 130 135 140

Tyr His

145

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

Ile

```
(vii)
         PRIOR APPLICATION DATA:
         (A) APPLICATION NUMBER: US 219,279
         (B) FILING DATE: 15-JUL-1988
(viii)
         ATTORNEY/AGENT INFORMATION:
         (A) NAME: Halluin, Albert P.
       (B) REGISTRATION NUMBER: 28,957
         (C) REFERENCE/DOCKET NUMBER: BIOG-20121
USA
         TELECOMMUNICATION INFORMATION:
 (ix)
         (A) TELEPHONE: 415-433-4150
          (B) TELEFAX: 415-433-8716
(2) INFORMATION FOR SEQ ID NO: 1:
  (i)
         SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 4 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear
 (ii)
         MOLECULE TYPE: peptide
(iii)
         HYPOTHETICAL: NO
 (iv)
         ANTI-SENSE: NO
 (xi)
         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
     Pro Xaa Gly Pro
     1
(2) INFORMATION FOR SEQ ID NO: 2:
  (i)
         SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 13 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear
 (ii)
         MOLECULE TYPE: DNA (genomic)
(iii)
         HYPOTHETICAL: NO
 (iv)
         ANTI-SENSE: NO
 (xi)
         SEQUENCE DESCRIPTION: SEQ ID NO: 2:
GGGTACCTGG GCC
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 886 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(ii)	MOLECUI	LE TYPE: DN/	A (genomi	ic)				
(iii)	нүротні	ETICAL: NO						
(iv)	ANTI-SI	ENSE: NO						
(vi)		AL SOURCE: RGANISM: Ch	inese cua	cumber				
(vii)		ATE SOURCE: LONE: alpha	-trichosa	anthin				
(ix)	(B) L(AME/KEY: CDS OCATION: 8.	.877					
(xi)	SEQUEN	CE DESCRIPT	ION: SEQ	ID NO: 3	5:			
CTCGAGG	ATG ATC	AGA TTC TTA	GTC CTC	TCT TTG	CTA ATT	CTC ACC	СТС	49
	Met Ile /	Arg Phe Leu 5	Val Leu	Ser Leu	Leu Ile 10	Leu Thr	Leu	
TTC CTA	ACA ACT	CCT GCT GTG	GAG GGC	GAT GTT	AGC TTC	CGT TTA	TCA	97
Phe Leu 15	Thr Thr I	Pro Ala Val 20	Glu Gly	Asp Val 25	Ser Phe	Arg Leu	Ser 30	
GGT GCA	ACA AGC	AGT TCC TAT	GGA GTT	TTC ATT	TCA AAT	CTG AGA	AAA 1	45
Gly Ala	Thr Ser S	Ser Ser Tyr 35	Gly Val	Phe Ile 40	Ser Asn	Leu Arg 45	Lys	
GCT CTT	CCA AAT	GAA AGG AAA	CTG TAC	GAT ATC	CCT CTG	TTA CGT	TCC 1	93
Ala Leu	Pro Asn (50	Glu Arg Lys	Leu Tyr 55	Asp Ile	Pro Leu	Leu Arg 60	Ser	
TCT CTT	CCA GGT	TCT CAA CGC	TAC GCA	TTG ATC	CAT CTC	ACA AAT	TAC 2	41
Ser Leu	Pro Gly 9	Ser Gln Arg	Tyr Ala 70	Leu Ile	His Leu 75	Thr Asn	Tyr	
GCC GAT	GAA ACC	ATT TCA GTG	GCC ATA	GAC GTA	ACG AAC	GTC TAT	ATT 2	89
Ala Asp 80	Glu Thr	Ile Ser Val 85	Ala Ile	Asp Val	Thr Asn 90	Val Tyr	Ile	
ATG GGA	TAT CGC	GCT GGC GAT	ACA TCC	TAT TTT	TTC AAC	GAG GCT	тст 3	37
Met Gly 95	Tyr Arg /	Ala Gly Asp 100	Thr Ser	Tyr Phe 105	Phe Asn	Glu Ala	Ser 110	
GCA ACA	GAA GCT	GCA AAA TAT	GTA TTC	AAA GAC	GCT ATG	CGA AAA	GTT 3	85
Ala Thr		Ala Lys Tyr 115	Val Phe	Lys Asp 120	Ala Met	Arg Lys 125	Val	

(D) TOPOLOGY: linear

ACG	CTT	CCA	TAT	TCT	GGC	AAT	TAC	GAA	AGG	стт	CAA	ACT	GCT	GCG	GGC	433
Thr	Leu	Pro	Туг 130	Ser	Gly	Asn	Туг	Glu 135	Arg	Leu	Gln	Thr	Ala 140	Ala	Gly	
AAA	ATA	AGG	GAA	AAT	ATT	CCG	стт	GGA	СТС	CCA	GCT	TTG	GAC	AGT	GCC	48
Lys	Ile	-Arg 145	Glu	Asn	Ile	Pro	Leu 150	Gly	Leu	Pro	Ala	Leu 155	Asp	Ser	Ala	
ATT	ACC	ACT	TTG	TTT	TAC	TAC	AAC	GCC	AAT	TCT	GCT	GCG	TCG	GCA	CTT	529
Ile	Thr 160	Thr	Leu	Phe	Туг	Tyr 165	Asn	Ala	Asn	Ser	Ala 170	Ala	Ser	Ala	Leu	
ATG	GTA	СТС	ATT	CAG	TCG	ACG	TCT	GAG	GCT	GCG	AGG	TAT	AAA	TTT	ATT	57
Met 175	Val	Leu	Ile	Gln	Ser 180	Thr	Ser	Glu	Ala 185	Ala	Arg	Туг	Lys	Phe	Ile 190	
GAG	CAA	CAA	ATT	GGG	AAG	CGC	GTT	GAC	AAA	ACC	TTC	CTA	CCA	AGT	TTA	62!
Glu	Gln	Gln	Ile	Gly 195	Lys	Arg	Val	Asp	Lys 200	Thr	Phe	Leu	Pro	Ser 205	Leu	
GCA	ATT	ATA	AGT	TTG	GAA	AAT	AGT	TGG	TCT	GCT	СТС	TCC	AAG	CAA	ATT	67.
Ala	Ile	Ile	Ser 210	Leu	Glu	Asn	Ser	Trp 215	Ser	Ala	Leu	Ser	Lys 220	Gln	Ile	
CAG	ATA	GCG	AGT	ACT	AAT	AAT	GGA	CAG	TTT	GAA	ACT	ССТ	GTT	GTG	CTT	72
Gln	Ile	Ala 225	Ser	Thr	Asn	Asn	Gly 230	Gln	Phe	Glu	Thr	Pro 235	Val	Val	Leu	
ATA	AAT	GCT	CAA	AAC	CAA	CGA	GTC	ATG	ATA	ACC	AAT	GTT	GAT	GCT	GGA	769
Ile	Asn 240	Ala	Gln	Asn	Gln	Arg 245	Val	Met	Ile	Thr	Asn 250	Val	Asp	Ala	Gly	•
GTT	GTA	ACC	TCC	AAC	ATC	GCG	TTG	CTG	CTG	AAT	CGA	AAC	AAT	ATG	GCA	817
Val 255	Val	Thr	Ser	Asn	Ile 260	Ala	Leu	Leu	Leu	Asn 265	Arg	Asn	Asn	Met	Ala 270	
GCC	ATG	GAT	GAC	GAT	GTT	ССТ	ATG	ACA	CAG	AGC	TTT	GGA	TGT	GGA	AGT	86!
Ala	Met	Asp	Asp	Asp 275	Val	Pro	Met	Thr	Gln 280	Ser	Phe	Gly	Cys	Gly 285	Ser	
TAT	GCT	ATT	TAG	TAACT	rcg /	A G										886
Tyr	Ala	Ile	290													

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 amino acids

										, `(
		-	-		emi										
(ii	i)	MC	DLECU	JLE 1	TYPE:	: pro	oteir	1							
(xi	i) '	Si	EQUE	NCE (DESC	RIPT	ION:	SEQ	ID I	NO:4:	:				
Met 1	Ile	Arg	Phe	Leu 5	Val	Leu	Ser	Leu	Leu 10	Ile	Leu	Thr	Leu	Phe 15	Leu
Thr	Thr	Pro	Ala 20	Val	Glu	Gly	Asp	Val 25	Ser	Phe	Arg	Leu	Ser 30	Gly	Ala
Thr	Ser	.Şer 35	Ser	Туг	Gly	Val	Phe	Ile 0	Ser	Asn	Ľeu	Arg 45	Lys	Ala	Leu
Pro	Asn 50	Glu	Arg	Lys	Leu	Tyr 55	Asp	Ile	Pro	Leu	Leu 60	Arg	Ser	Ser	Leu
Pro 65	Gly	Ser	Gln	Arg	Туг 70	Ala	Leu	Ile	His	Leu 75	Thr	Asn	Туг	Ala	Asp 80
Glu	Thr	Ile	Ser	Val 85	Ala	Ile	Asp	Val	Thr 90	Asn	Val	Туг	Ile	Met 95	Gly
Tyr	Arg	Ala	Gly 100	Asp	Thr	Ser	Туг	Phe 105	Phe	Asn	Glu	Ala	Ser 110	Ala	Thr
Glu	Ala	Ala 115	Lys	Туг	Val	Phe	Lys 120	Asp	Ala	Met	Arg	Lys 125	Val	Thr	Leu
Pro	Tyr 130	Ser	Gly	Asn	Туг	Glu 135	Arg	Leu	Gln	Thr	Ala 140	Ala	Gly	Lys	Ιle
Arg 145	Glu	Asn	Ile	Pro	Leu 150	Gly	Leu	Pro	Ala	Leu 155	Asp	Ser	Ala	Ile	Thr 160
Thr	Leu	Phe	Tyr	Tyr 165	Asn	Ala	Asn	Ser	Ala 170	Ala	Ser	Ala	Leu	Met 175	Val
Leu	Ile	Gln	Ser 180	Thr	Ser	Glu	Ala	Ala 185	Arg	Туг	Lys	Phe	Ile 190	Glu	Glr
Gln	Ile	Gly 195	Lys	Arg	Val	Asp	Lys 200	Thr	Phe	Leu	Pro	Ser 205	Leu	Ala	Ile
Ile	Ser 210	Leu	Glu	Asn	Ser	Trp 215	Ser	Ala	Leu	Ser	Lys 220	Gln	Ile	Gln	Ile
Ala 225	Ser	Thr	Asn	Asn	Gly 230	Gln	Phe	Glu	Thr	Pro 235	Val	Val	Leu	Ile	Asr 240
Ala	Gln	Asn	Gln	Arg 245	Val	Met	Ile	Thr	Asn 250	Val	Asp	Ala	Gly	Val 255	Val
Thr	Ser	Asn	Ile	Ala	Leu	Leu	Leu	Asn	Arg	Asn	Asn	Met	Ala	Ala	Met

. 265

Asp Asp Asp Val Pro Met Thr Gln Ser Phe Gly Cys Gly Ser Tyr Ala

(2)	IN	FORM/	ATIO	N FOI	R SEG	ID	NO:	5:								
(1	i)	• (/	4) l	LENG		1450	base	e pa	irs							
					: nuc											
					OGY:			igic								
(i	i)	M	DLECI	JLE .	TYPE:	: DN	A (g	enom [*]	ic)							
(ii	i)	н	POTI	HETI	CAL:	NO										
(iv	/)	AI	NT I - S	SENSI	E: NO)										
(v	i)				SOUR											
		G	A) (ORGAI	II SM:	: Ory	yza :	sativ	va							
(vi	i)				SOUF		- amv	معد ا								
		`	.,			· p···u	u ,									
(i)	()		EATU							4						
					TION:) LO(316	CATIO	JN:	12.	. 1510	•			
(x	i)	SI	EQUE	NCE (ESCF	RIPT	ION:	SEQ	ID I	NO: 5	5:					
ССТ	CGAG	STG (CATO	G CA	G GT(G CT	G AA	C AC	C AT	G GTC	G AA	CA	CAC '	TTC '	TTG	48
				t Gli 1	n Val	l Lei		n Thi	r Me	t Val	l Ası	n Ly:		s Pho	e Leu	ı
TCC	стт	TCG	GTC	стс	ATC	GTC	СТС	стт	GGC	стс	TCC	TCC	AAC	TTG	ACA	96
Ser	Leu 15	Ser	Val	Leu `.	Ile	Val 20	Leu	Leu	Gly	Leu	Ser 25	Ser	Asn	Leu	Thr	
GCC	GGG	CAA	GTC	CTG	TTT	CAG	GGA	TTC	AAC	TGG	GAG	TCG	TGG	AAG	GAG	144
Ala 30	Gly	Gln	Val	Leu	Phe 35	Gln	Gly	Phe	Asn	Тгр 40	Glu	Ser	Тгр	Lys	Glu 45	
AAT	GGC	GGG	TGG	TAC	AAC	TTC	CTG	ATG	GGC	AAG	GTG	GAC	GAC	ATC	GCC	192
Asn	Gly	Gly	Trp	Tyr 50	Asn	Phe	Leu	Met	Gly 55	Lys	Val	Asp	Asp	Ile 60	Ala	
GCA	GCC	GGC	ATC	ACC	CAC	GTC	TGG	стс	ССТ	CCG	CCG	TCT	CAC	тст	GTC	240
Ala	Ala	Gly	Ile 65	Thr	His	Val	Trp	Leu 70	Pro	Pro	Pro	Ser	His 75	Ser	Val	
GGC	GAG	CAA	GGC	TAC	ATG	CCT	GGG	CGG	CTG	TAC	GAT	CTG	GAC	GCG	тст	288

Gly Glu Gln Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser

90

85

. 80

AAG TAC GGC AAC GAG GCG CAG CTC AAG TCG CTG ATC GAG GCG TTC CA	AT 336
Lys Tyr Gly Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe Hi 95 100 105	is
GGC AAG GGC GTC CAG GTG ATC GCC GAC ATC GTC ATC AAC CAC CGC AC	CG 384
Gly Lys∲Gly Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Th 110 115 120 17	nr 25
GCG GAG CAC AAG GAC GGC CGC GGC ATC TAC TGC CTC TTC GAG GGC GC	GG 432
Ala Glu His Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gl 130 135 140	
ACG CCC GAC TCC CGC CTC GAC TGG GGC CCG CAC ATG ATC TGC CGC GA	AC 480
Thr Pro Asp Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg As 145 150 155	sp
GAC CCC TAC GGC CAT GGC ACC GGC AAC CCG GAC ACC GGC GCC GAC TO	rc 528
Asp Pro Tyr Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Ph 160 165 170	ne .
GCC GCC GCG CCG GAC ATC GAC CAC CTC AAC AAG CGC GTC CAG CGG G/	AG 576
Ala Ala Ala Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Gl 175 180 185	lu
CTC ATT GGC TGG CTC GAC TGG CTC AAG ATG GAC ATC GGC TTC GAC GC	CG 624
Leu Ile Gly Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Al 190 195 200 20	
TGG CGC CTC GAC TTC GCC AAG GGC TAC TCC GCC GAC ATG GCA AAC A'	rc 672
Trp Arg Leu Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys II 210 215 220	le
210 215 220	CG 720
TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG ACC TYR Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Th	CG 720
TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG AGC Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Th 225 230 235	cg 720 nr cg 768
TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG AGC Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Th 225 230 235 TCC ATG GCG AAC GGC GGG GAC GGC AAG CCG AAC TAC GAC CAG AAC GG Ser Met Ala Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn A	CG 720 CG 768
TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG AG Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Th 225 TCC ATG GCG AAC GGC GGG GAC GGC AAG CCG AAC TAC GAC CAG AAC GG Ser Met Ala Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Al 240 225 220 220 221 220 221 222 223 225 226	CG 720 CG 768 La AC 816
TAC ATC GAC GCC ACC GAG CCG AGC TTC GCC GTG CCC GAG ATA TCG AGC Tyr Ile Asp Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp The 225 230 235 TCC ATG GCG AAC GGC GGG GAC GGC AAG CCG AAC TAC GAC CAG AAC GGC Ser Met Ala Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Al 240 245 250 CAC CGG CAG GAG CTG GTC AAC TGG GTC GAT CGT GTC GGC GGC GCC AAC His Arg Gln Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asp	CG 720 Or 768 La AC 816

										- 1						
GCC	GTG	GAG	GGC	GAG	CTG	TGG	CGC	CTC	CGC	GGC	GAG	GAC	GGC	AAG	GCG	912
Ala	Val	Glu	Gly	Glu 290	Leu	Trp	Arg	Leu	Arg 295	Gly	Glu	Asp	Gly	Lys 300	Ala	
ccc	GGC	ATG	ATC	GGG	TGC	TGG	CCG	GCC	AAG	GCG	ACG	ACC	TTC	GTC	GAC	960
Pro	Gly	Met	I l e 305	Gly	Trp	Тгр	Pro	Ala 310	Lys	Ala	Thr	Thr	Phe 315	Val	Asp	
AAC	CAC	GAC	ACC	GGC	TCG	ACG	CAG	CAC	CTG	TGG	CCG	TTC	ccc	TCC	GAC	1008
Asn	His	Asp 320	Thr	Gly	Ser	Thr	Gln 325	His	Leu	Trp	Pro	Phe 330	Pro	Ser	Asp	
AAG	GTC	ATG	CAG	GGC	TAC	GCA	TAC	ATC	стс	ACC	CAC	ССС	GGC	AAC	CCA	1056
Lys	Val 335	Met	Gln	Gly	Tyr	Ala 340	Tyr	Ile	Leu	Thr	His 345	Pro	Gly	Asn	Pro	
TGC	ATC	TTG	TAC	GAC	CAT	TTC	TTC	GAT	TGG	GGT	стс	AAG	GAG	GAG	ATC	1104
Cys 350	Ile	Phe	Tyr	Asp	His 355	Phe	Phe	Asp	Тгр	Gly 36 0	Leu	Lys	Glu	Glu	I l e 365	•
GAG	CGC	CTG	GTG	TCA	ATC	AGA	AAC	CGG	CAG	GGG	ATC	CAC	CCG	GCG	AGC	1152
Glu	Arg	Leu	Val	Ser 370	Ile	Arg	Asn	Arg	Gln 375	Gly	Ile	His	Pro	Ala 380	Ser	
GAG	CTG	CGC	ATC	ATG	GAA	GCT	GAC	AGC	GAT	СТС	TAC	стс	GCG	GAG	ATC	1200
Glu	Leu	Arg	Ile 385	Met	Glu	Ala	Asp	Ser 390	Asp	Leu	Tyr	Leu	Ala 395	Glu	Ile	
GAT	GGC	AAG	GTG	ATC	ACA	AAG	ATT	GGA	CCA	AGA	TAC	GAC	GTC	GAA	CAC	1248
Asp	Gly	Lys 400	Val	Ile	Thr	Lys	Ile 405	Gly	Pro	Arg	Tyr	Asp 410	Val	Glu	His	
стс	ATC	CCC	GAA	GGC	TTC	CAG	GTC	GTC	GCG	CAC	GGT	GAT	GGC	TAC	GCA	1296
Leu	I l e 415	Pro	Glu	Gly	Phe	Gln 420	Val	Val	Ala	His	Gly 425	Asp	Gly	Tyr	Ala	
ATC	TGG	GAG	AAA	ATC	TGAC	GCGC/	ACG /	ATGA	CGAG/	AC TO	CTCA	STTT	A GC	AGAT	ГТАА	1351
Ile 430	Trp	Glu	Lys	Lle	435	5										
CCT	CGAT	ITT 1	TAC	CTG	AC CO	GTAT	TACG	r at/	ATAC	STGC	CGG	CAAC	GAG (CTGT	ATCCG	A 1411
TCC	AATI	rac (GATO	CAA	IT G1	CCAC	GAA	G TC	CTCG/	AGG						1450

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 434 amino acids (B) TYPE: amino acid													
(D) Topology: linear													
(ii) . MOLECULE TYPE: protein													
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:													
Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser 1 5 10 15													
Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr Ala Gly Gln 20 25 30													
Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu Asn Gly Gly 35 40 45													
Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala Ala Gly 50 55 60													
Ile Thr His Val Trp Leu Pro Pro Pro Ser His Ser Val Gly Glu Gln 65 70 75 80													
Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser Lys Tyr Gly 85 90 95													
Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe His Gly Lys Gly 100 105 110													
Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Thr Ala Glu His 115 120 125													
Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gly Thr Pro Asp 130 135 140													
Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg Asp Asp Pro Tyr 145 150 155 160													
Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Phe Ala Ala Ala 165 170 175													
Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Glu Leu Ile Gly 180 185 190													
Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Ala Trp Arg Leu 195 200 205													
Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys Ile Tyr Ile Asp 210 215 220													
Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Thr Ser Met Ala 225 230 235 240													
Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Ala His Arg Gln 245 250 255													
Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asn Ser Asn Gly 260 265 270													

Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val Ala Val Glu

Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala Pro Gly Met 290 295 300

Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp Asn His Asp 305 310 315 320

Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp Lys Val Met 325 330 335

Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro Cys Ile Phe 340 345 350

Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile Glu Arg Leu 355 360 365

Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser Glu Leu Arg 370 375 380

Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile Asp Gly Lys 385 390 395 400

Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His Leu Ile Pro 405 410 415

Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala Ile Trp Glu 420 425 430

Lys Ile

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 709 base pairs

(B) TYPE: nucleic acid

(G) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:

(B) CLONE: alpha-hemoglobin

(ix) FEATURE:

(A) NAME/KEY: transit_peptide (B)

LOCATION: 26. .241

(B) LOCATION: 26. .241

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 245. .670

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CTCGAGGGCA TCT	GATCTTT CAAGA	ATGGC ACAAA	TTAAC AACATGGCAC	AAGGGATACA 60							
AACCCTTAAT CCC	AATTCCA ATTTC	CATAA ACCCC	AAGTT CCTAAATCTT	CAAGTTTTCT 120							
TGTTTTT GA TGT	AAAAAAC TGAAA	AATTC AGCAA	ATTCT ATGTTGGTTT	TGAAAAAAGA 18							
TTCAATTTTT ATG	CAAAAGT TTTGT	TCCTT TAGGA	TTTCA GCAGGTGGTA	GAGTTTCTTG 240							
CATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGC											
Val Leu S 1	er Pro Ala As 5	p Lys Thr A	sn Val Lys Ala A 10	la Trp Cly 15							
AAG GTT GGC GC	G CAC GCT GGC	GAG TAT GG	T GCG GAG GCC CT	G GAG AGG 337							
Lys Val Gly Al	a His Ala Gly 20	Glu Tyr Gl 2	y Ala Glu Ala Le 5	u Glu Arg 30							
ATG TTC CTG TC	C TTC CCC ACC	ACC AAG AC	C TAC TTC CCG CA	C TTC GAC 385							
Met Phe Leu Se		Thr Lys Th 40	r Tyr Phe Pro Hi: 4								
CTG AGC CAC GG	C TCT GCC CAG	GTT AAG GG	C CAC GGC AAG AA	G GTG GCC 433							
Leu Ser His Gl	y Ser Ala Glr	Val Lys Gl 55	y His Gly Lys Ly: 60	s Val Ala							
GAC GCG CTG AC	C AAC GCC GTG	GCG CAC GT	G GAC GAC ATG CC	C AAC GCG 481							
Asp Ala Leu Th 65	r Asn Ala Val 70		l Asp Asp Met Pro 75	o Asn Ala							
CTG TCC GCC CT	G AGC GAC CTG	CAC GCG CA	C AAG CTT CGG GT	G GAC CCG 529							
Leu Ser Ala Le 80	u Ser Asp Leu 85	His Ala Hi	s Lys Leu Arg Va 90	l Asp Pro 95							
GTC AAC TTC AA	G CTC CTA AGO	CAC TGC CT	G CTG GTG ACC CT	G GCC GCC 577							
Val Asn Phe Ly	s Leu Leu Ser 100	His Cys Le	u Leu Val Thr Le 5	JAla Ala 110							
CAC CTC CCC GC	C GAG TTC ACC	CCT GCG GT	G CAC GCC TCC CT	G GAC AAG 625							
His Leu Pro Al		Pro Ala Va 120	l His Ala Ser Le 12	• •							
TTC CTG GCT TC	T GTG AGC ACC	GTG CTG AC	C TCC AAA TAC CG	T TAAGCTGGAG 677							
Phe Leu Ala Se 130	r Val Ser Thr	Val Leu Th 135	r Ser Lys Tyr Ar 140	9							

2) I	NFOR	RMATI	ON FO	R SE	Q ID	NO:	В:			.				
(i)	л	SEQUI (A) (B) (D)	LENO TYPE	iTH: : a	141 mino	am aci	ino d	acid	s					
(ii)	ø	MOLE	CULE	TYPE	: pr	otei	n							
(ix)		SEQU	ENCE	DESC	DIDT	ION•	SEU	in :	NO.8					
al Le											Δla	Ten	c l v	lve
1	u 30	SI F1	5		Lys		ASII	10		ALG	Αια		15	
al Gl	y Al	la Hi 2		Gly	Glu	Туг	Gly 25		Glu	Ala	Leu	Glu 30		Met
he Le		er Ph	e Pro) Thr	Thr	Lys 40		Туг	Phe	Pro	His 45	Phe	Asp	Leu
Ger Hi 5		ly Se	r Ala	Gln	Val 55		Gly	His	Gly	Lys 60		Val	Ala	Asp
ila Le 65	u Th	hr As	n Ala	70		His	Val	Asp	Asp 75		Pro	Asn	Ala	Leu 80
er Al	a Le	eu Se	r Asp 85		His	Ala	His	Lys 90		Arg	Val	Asp	Pro 95	
sn Ph	e Ly	ys Le 10		ı Ser	His	Cys	Leu 105		Val	Thr	Leu	Ala 110		His
eu Pr	o Al 11		u Phe	e Thr	Pro	Ala 120		His	Ala	Ser	Leu 125	Asp	Lys	Phe
eu Al 13		er Va	l Ser	Thr	Val 135		Thr	Ser	Lys	Tyr 140				
2) _, IN	FORM	OITAM	N FOR	SEQ	ID :	NO:9	:							
(i)		SEQU (A) (B) (C) (D)	LENGT Type: Stran	H: 7 nuc	43 b leic ESS:	ase aci sin	pair d	s						
(ii)		MOLE	CULE	TYPE	: cD	NA t	o mR	NA						
iii)		нүро	THETI	CAL:	NO									
(iv)		ANTI	SENS	E: N	0									
(vi)		ORIG (A)				mo s	apie	ns						
vii)		IMME				hemo	glob	in						

.

LOCATION: 26241												
(B) LOCATION: 26241												
(ix) FEATURE:												
(A) NAME/KEY: CDS												
(B) LOCATION: 245685												
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:												
CTCGAGGGGA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC AAGGGATACA												
AACCCTTAAT CCCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT CAAGTTTTCT 12												
TGTTTTTGGA TCTAAAAAAC TGAAAAATTC AGCAAATTCT ATGTTGGTTT TGAAAAAAAGA 18												
TICAATITIT ATGCAAAAGT TITGTTCCTT TAGGATTTCA GCAGGTGGTA GAGTTTCTTG	240											
GATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG	289											
Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp 1 5 10 15												
1 3 10 13												
GGC AAG GTG AAC GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG	337											
Gly Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu 20 25 30												
CTG GTG GTC TAC CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAT	385											
Leu Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp 35 40 45												
CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAT	433											
Leu Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His 50 55 60												
GGC AAG AAA GTG CTG GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC	481											
Gly Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp 65 70 75												
AAC CTC AAG GGC ACC TTT GCC ACC CTG AGT GAG CTG CAC TGT GAC AAG	529											
Asn Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys 80 85 90 95												
CTG CAC GTG GAT CCT GAG AGC TTC AGG CTC CTA GGC AAC GTG CTG GTC	577											
Leu His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val												
TGT GTG CTG GCG CAT CAC TIT GGC AAA GAA TTC ACC CCA CCA GTG CAG	625											
Cys Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln 115 120 125	٠											
GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC CAC	673											

(ix)

FEATURE:

(A) NAME/KEY: transit_peptide (B)

										•	_				
Ala	Ala	Tyr 130	Gln	Lys	Val	Val	Ala 135	Gly	Val	Ala	Asn	Ala 140	Leu	Ala	His
AAG	AG TAT, CAC TAAGCTCGCT TICTIGCTGT CCAATITCTA TTAAAGGTTC														
Lys	.ys Tyr His 145 [™]														
CTTI	CTTTGTGGGG TCGAGGTCGA C														
(2)	(2) INFORMATION FOR SEQ ID NO: 10:														
'(i	(i) SEQUENCE CHARACTERISTICS:														
	(A) LENGTH: 146 amino acids														
	(B) TYPE: amino acid (D) TOPOLOGY: linear														
(ii	i)	•	•				oteir	1							
(xi	i)	SE	QUEN	ICE C	ESCR	IPT I	ON:	SEQ	ID N	i O: 1	10:				
Val 1	His	Leu	Thr	Pro 5	Glu	Glu	Lys	Ser	Ala 10	Val	Thr	Ala	Leu	Trp 15	Gly
Lys	Val	Asn	Val 20	Asp	Glu	Val	Gly	Gly 25	Glu	Ala	Leu	Gly	Arg 30	Leu	Leu
Val	Val	Tyr 35	Pro	Trp	Thr	Gln	Arg 40	Phe	Phe	Glu	Ser	Phe 45	Gly	Asp	Leu
Ser	Thr 50	Pro	Asp	Ala	Val	Met 55	Gly	Asn	Pro	Lys	Val 60	Lys	Ala	His	Gly
Lys 65	Lys	Val	Leu	Gly	Ala 70	Phe	Ser	Asp	Gly	Leu 75	Ala	His	Leu	Asp	Asn 80
Leu	Lys	Gly	Thr	Phe 85	Ala	Thr	Leu	Ser	Glu 90	Leu	His	Cys	Asp	Lys 95	Leu
His	Val	Asp	Pro 100	Glu	Ser	Phe	Arg	Leu 105	Leu	Gly	Asn	Val	Leu 110	Val	Cys
Val	Leu	Ala 115	His	His	Phe	Gly	Lys 120	Glu	Phe	Thr	Pro	Pro 125	Val	Gln	Ala
Ala	Tyr 130	Gln	Lys	Val	Val	Ala 135	Gly	Val	Ala	Asn	Ala 140	Leu	Ala	His	Lys
Tyr 145	His														
(2)	INI	FORMA	ATION	l FOF	R SEG	ID	NO:	11:							

(i)

SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

722

743

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SCURCE:

(A) ORGANISM: alkalophilic Bacillus sp.

(B) STRAIN: 38-2

(vii) IMMEDIATE SCURCE:

(B) CLONE: beta-cyclodextrin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Ala Pro Asp Thr Ser Val Ser Asn Lys Gln Asn Phe Ser Thr Asp Val

1 5 10 15

Ile

(D)	TOPOLOGY:	linear	1
		•	•

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe Leu
1 5 10 15

Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser Gly Ala 20 25 30

Thr Ser Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys Ala Leu 35 40 45

Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser Ser Leu
50 55 60

Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr Ala Asp 65 70 75 80

Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile Met Gly
85 90 95

Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser Ala Thr 100 105 110

Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val Thr Leu 115 120 125

Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly Lys Ile 130 · 135 140

Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala Ile Thr 145 150 155 160

Thr Leu Phe Tyr Tyr Asn Ala Asn Ser Ala Ala Ser Ala Leu Met Val
165 170 175

Leu Ile Gln Ser Thr Ser Glu Ala Ala Arg Tyr Lys Phe Ile Glu Gln 180 185 190

Gln Ile Gly Lys Arg Val Asp Lys Thr Phe Leu Pro Ser Leu Ala Ile 195 200 205

Ile Ser Leu Glu Asn Ser Trp Ser Ala Leu Ser Lys Gln Ile Gln Ile 210 215 220

Ala Ser Thr Asn Asn Gly Gln Phe Glu Thr Pro Val Val Leu Ile Asn 225 230 235 240

Ala Gln Asn Gln Arg Val Met Ile Thr Asn Val Asp Ala Gly Val Val
245 250 255

Thr Ser Asn Ile Ala Leu Leu Leu Asn Arg Asn Asn Met Ala Ala Met 260 265 270

Asp Asp Asp Val Pro Met Thr Gln Ser Phe Gly Cys Gly Ser Tyr Ala 275 280 285

2)	INI	ORM/	AT I OI	FOR	SEC) ID	NO:	5:		٠							
c	i)	SI	QUE	ICE (HARA	CTER	RIST	ICS:	•								
•	•			.ENG1					irs								
				YPE:				•									
				STRAN													
				TOPOL													
(ii	i)	MC	DLECU	JLE 1	YPE:	: DN/	\ (g	enom [*]	ic)								
iii	i)	н	POTI	HETIC	CAL:	NO											
(iv	/)	Al	NTI-S	SENSE	: NC)											
(vi	i)	O	RIGII	IAL S	OUR	Œ:											
		(/	A) (ORGAN	IISM:	Ory	/za s	sativ	∨a								
vii	i)	11	MED 1	ATE	SOUF	RCE:											
		(I	3) (CLONE	: al	lpha-	ату	lase		•							
(i)	()	F	EATU	RE:					•								
		(/	4) !	NAME/	KEY:	CDS	(B)) LO	CATIO	ON: 1	12.	. 1316	5				
		(1	3) (_OCA1	ION:	12.	13	316									
(xi	i)	SI	EQUE	CE C	ESCF	RIPTI	ON:	SEQ	ID I	10: 5	5 :						
CTO	CGAG	STG (CATO	G CAC	GTO	СТС	AA G	CAC	CATO	G _, GTC	AA(CAC	CAC 1	TTC 1	TTG	48	
			Met	t Glr	n Val	Leu	ı Ası	n Thi	r Mei	t Val	. Ası	n Lys	s His	s Phe	e Lei	и	
				١			!					10					
CC	СТТ	TCG	GTC	СТС	ATC	GTC	СТС	CTT	GGC	СТС	TCC	TCC	AAC	TTG	ACA	96	
er	l eu	Ser	Val	l eu	Ile	Val	Leu	Leu	Gly	l eu	Ser	Ser	4en	l eu	Thr		
	15	-			•••	20			J.,	200	25	00.	AUIT		••••		
CC	GGG	CAA	GTC	CTG	TTT	CAG	GGA	ттс	AAC	TGG	GAG	TCG	TGG	AAG	GAG	144	
la	Gly	Gln	Val	Leu	Phe	Gln	Gly	Phe	Asn	Trp	Glu	Ser	Trp	Lvs	Glu		
30					35					40				-,-	45		
ΑT	GGC	GGG	TGG	TAC	AAC	TTC	CTG	ATG	GGC	AAG	GTG	GAC	GAC	ATC	GCC	192	
	Clv	Clv	T	Tun	A am	Dha.	1	W.A	C1	1	V-1		4			•	
SN	ыцу	цц	ırp	50	ASN	Pne	Leu	мет	Gly 55	-	val	ASP	ASP	60	ALB		
·CA	ccc	ccr	ATC	۸۲۲	CAC	CTC	TCC	CTC	CCT	ccc	ccc	TCT	CAC	TCT	CTC	240	
CA	ucc	duc	AIC	ACC	CAC	GIC	100		CCI	CCG	CCG	161	LAL	161	616	240	
la	Ala	Gly	Ile 65	Thr	His	Val	Trp	Leu 70	Pro	Pro	Pro	Ser	His 75	Ser	Val		
			-					, 0					, ,			*	
GC	GAG	CAA	GGC	TAC	ATG	CCT	GGG	CGG	CTG	TAC	GAT	CTG	GAC	GCG	TCT	288	
ly	Glu		Gly	Tyr	Met	Pro		Arg	Leu'	Tyr	Asp		Asp	Ala	Ser		
		80					85					90					

AAG TAC GGC AAC GAG GCG CAG CTC AAG T	CG CTS ATC GAG GCG TTC CAT	336
Lys Tyr Gly Asn Glu Ala Gln Leu Lys S 95 100	er Leu Ile Glu Ala Phe His 105	
GGC AAG GGC GTC CAG GTG ATC GCC GAC A	TC GTC ATC AAC CAC CGC ACG	384
Gly Lys Gly Val Gln Val Ile Ala Asp I 110 115	le Val Ile Asn His Arg Thr 120 125	
GCG GAG CAC AAG GAC GGC CGC GGC ATC T	AC TGC CTC TTC GAG GGC GGG	432
Ala Glu His Lys Asp Gly Arg Gly Ile T 130 1	yr Cys Leu Phe Glu Gly Gly 35 140	
ACG CCC GAC TCC CGC CTC GAC TGG GGC C	CG CAC ATG ATC TGC CGC GAC	480
Thr Pro Asp Ser Arg Leu Asp Trp Gly P 145 150	ro His Met Ile Cys Arg Asp 155	
GAC CCC TAC GGC CAT GGC ACC GGC AAC C	CG GAC ACC GGC GCC GAC TTC	528
Asp Pro Tyr Gly Asp Gly Thr Gly Asn P 160 165	ro Asp Thr Gly Ala Asp Phe 170	
GCC GCC GCG CCG GAC ATC GAC CAC CTC A	AC AAG CGC GTC CAG CGG GAG	576
Ala Ala Ala Pro Asp Ile Asp His Leu A 175 180	sn Lys Arg Val Gln Arg Glu 185	
CTC ATT GGC TGG CTC GAC TGG CTC AAG A	TG GAC ATC GGC TTC GAC GCG	624
Leu Ile Gly Trp Leu Asp Trp Leu Lys M 190 195	et Asp Ile Gly Phe Asp Ala 200 205	
TGG CGC CTC GAC TTC GCC AAG GGC TAC T	CC GCC GAC ATG GCA AAC ATC	672
Trp Arg Leu Asp Phe Ala Lys Gly Tyr S 210 2	er Ala Asp Met Ala Lys Ile 15 220	
TAC ATC GAC GCC ACC GAG CCG AGC TTC G	CC GTG CCC GAG ATA TCG ACG	720
Tyr Ile Asp Ala Thr Glu Pro Ser Phe A 225 230	la Val Ala Glu Ile Trp Thr 235	
TCC ATG GCG AAC GGC GGG GAC GGC AAG C	CG AAC TAC GAC CAG AAC GCG	768
Ser Met Ala Asn Gly Gly Asp Gly Lys P 240 245	ro Asn Tyr Asp Gln Asn Ala 250	
CAC CGG CAG GAG CTG GTC AAC TGG GTC G	AT CGT GTC GGC GGC GCC AAC	816
His Arg Gln Glu Leu Val Asn Trp Val A 255 260	sp Arg Val Gly Gly Ala Asn 265	
ACC AAC GGC ACG GCG TTC GAC TTC ACC A	CC AAG GGC ATC CTC AAC GTC	, 864
Ser Asn Gly Thr Ala Phe Asp Phe Thr T 270 275	hr Lys Gly Ile Leu Asn Val 280 285	

GCC GTG GAG GG	C GAG CTG	TGG CGC	CTC CGC	GGC GAG	GAC GGC	AAG GCG	912
Ala Val Glu Gl	y Glu Leu 290	Trp Arg	Leu Arg 295	Gly Glu	Asp Gly	Lys Ala 300	
CCC GGC ATG AT	C GGG TGC	TGG CCG	GCC AAG	GCG ACG	ACC TTC	GTC GAC	960
Pro Gly Met Il 30		•	Ala Lys 310	Ala Thr	Thr Phe 315	Val Asp	
AAC CAC GAC AC	C GGC TCG	ACG CAG	CAC CTG	TGG CCG	TTC CCC	TCC GAC	1008
Asn His Asp Th 320	r Gly Ser	Thr Gln 325	His Leu	Trp Pro	Phe Pro 330	Ser Asp	
AAG GTC ATG CA	G GGC TAC	GCA TAC	ATC CTC	ACC CAC	CCC GGC	AAC CCA	1056
Lys Val Met Gl 335	n Gly Tyr	Ala Tyr 340	Ile Leu	Thr His 345	Pro Gly	Asn Pro	
TGC ATC TTG TA	C GAC CAT	TTC TTC	GAT TGG	GGT CTC	AAG GAG	GAG ATC	1104
Cys Ile Phe Ty 350	r Asp His 355	Phe Phe	Asp Trp	Gly Leu 360	Lys Glu	Glu Ile 365	
GAG CGC CTG GT	G TCA ATC	AGA AAC	CGG CAG	GGG ATC	CAC CCG	GCG AGC	1152
Glu Arg Leu Va	l Ser Ile 370	Arg Asn	Arg Gln 375	Gly Ile	His Pro	Ala Ser 380	
GAG CTG CGC AT	C ATG GAA	GCT GAC	AGC GAT	CTC TAC	CTC GCG	GAG ATC	1200
Glu Leu Arg Il 38		•	Ser Asp 390	Leu Tyr	Leu Ala 395	Glu Ile	
GAT GGC AAG GT	G ATC ACA	AAG ATT	GGA CCA	ÁGA TAC	GAC GTC	GAA CAC	1248
Asp Gly Lys Va 400	l Ile Thr	Lys Ile 405	Gly Pro	Arg Tyr	Asp Val 410	Glu His	
CTC ATC CCC GA	A GGC TTC	CAG GTC	GTC GCG	CAC GGT	GAT GGC	TAC GCA	1296
Leu Ile Pro Gl 415	u Gly Phe	Glm Val 420	Val Ala	His Gly 425	Asp Gly	Tyr Ala	
ATC TGG GAG AA	A ATC TGA	GCGCACG A	TGACGAGA	C TCTCAC	STTTA GC/	AGATTTAA	1351
Ile Trp Glu Ly 430	s LIe 43	5					
CCTGCGATTT TTA	CCCTGAC C	GGTATACGT	ATATACO	STGC CGG(CAACGAG (CTGTATCCGA	1411
TCCGAATTAC GGA	TGCAATT G	TCCACGAAG	TCCTCGA	NGG			1450

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS: